
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

Object Recognition for Continuous Drone Monitoring

[1] T. Vardhan , [2] N. Bhavani, [3] K. Kushi Preetham, [4] B. Sai Nithin, [5] G. Vaishnavi, [6] Mr. Mohd Nawazuddin, [7] Mrs. G. Suguna

[1,2,3,4,5]UG Scholar, Dept. of CS, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

[6] Assistant Professor, Dept. of CSE, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

[7] Assistant Professor, Dept. of EEE, Narsimha Reddy Engineering College, Maisammaguda, Kompally, Secunderabad, India.

***Abstract*:**

When it comes to analyzing photos taken by UAVs, deep learning (DL) has become an invaluable tool for remote sensing. This review seeks to provide a thorough grasp of the topic, even though it has produced notable contributions in other domains. Included in the paper is a comprehensive review of current methods and their real-world uses for object recognition in drone surveillance in real-time. Useful Search Terms: Unidentified Aerial Vehicles, Deep Learning, One-Stage Detector, Two-Stage Detector.

I. INTRODUCTION

Unmanned Aerial Vehicles, more often known as drones, have found several uses due to their capacity to reach inaccessible or hazardous locations. UAVs have cameras that can take video or stills from different angles and heights; they have many potential uses in fields as diverse as aerial photography, environmental monitoring, defense, and search and rescue. In real-time applications, manual tracking and picture collection are not feasible; hence, machine learning methods are extensively used to construct an automated system capable of processing and analyzing UAV-captured photos. However, it has nothing to do with the actual act of taking pictures, whether for mapping, surveying, surveillance, or inspection. The UAV has the capability to wirelessly send the photographs in real-time to a ground station or store them onboard for later retrieval and analysis. Figure 1 shows the fundamental design of the drone surveillance system. An onboard camera is usually the means by which images are captured in drone surveillance. Depending on the task at hand, the camera might use thermal or multispectral sensors instead of a regular RGB camera. While the drone is in the air, the camera records footage or stills of the ground below. Either the footage or stills will be sent live to a base station or kept on the drone for further review at a later time. The drone cameras might have their features and settings tweaked to provide the best possible images and to record certain kinds of data. You can control the camera's exposure, focus, and zoom, and it may even include modes for taking still photographs and video. If the drone has other sensors and technologies, such GPS and LIDAR, it may aid navigate and map the area being surveyed in addition to the camera.
Images taken in different locations will either be saved in batches or tracked in real-time. Object detection is the procedure that allows us to track certain objects. In drone surveillance, object identification is finding and locating certain things in the footage or stills taken by the aircraft. Agriculture, environmental monitoring, SAR, and military activities are just a few of the many fields that have found drone surveillance to be an indispensable tool. Drones are a popular alternative for surveillance activities due to their capacity to effectively and swiftly deliver overhead images of broad regions. Drone surveillance can only be as good as the object identification algorithms used to monitor them. Drone identification is more difficult than using fixed cameras for a number of reasons, including the fact that these UAVs are flown at great heights: From above, you can see everything clearly, but you'll also have to contend with perspective distortion, shadows, and reflections. Environment without control: Difficult elements, including climate, illumination, and environmental changes throughout time, things in motion: identifying and following things that could be traveling at great speeds or making sudden directional changes, The complexity and precision of object identification algorithms might be constrained by the limited processing power and memory of drones, in comparison to other computer systems.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

II. OBJECT DETECTION UAV OVERVIEW

An Object: What Is It? An item is anything that can be graphically represented by elements that are retrieved from UAVs. The process of object detection begins with the identification and localization of various elements, such as crops, flowers, people, weapons, etc., in order to offer information about the item's position or condition. In a nutshell, it classifies the extracted element. A development of ML, Deep Learning (DL) is based on the hierarchical organization of the human brain and its ability to solve problems. extensive and varied set of uses. Because DL architecture takes use of deeper combinations of input and hidden layers, it is able to analyze images and extract features from complicated and huge datasets more efficiently. Drone processing applications, where data diversity and difficulty in human processing are major concerns, greatly benefit from its high processing capabilities. This might be one explanation for the widespread use of DL in data-driven and image processing applications.

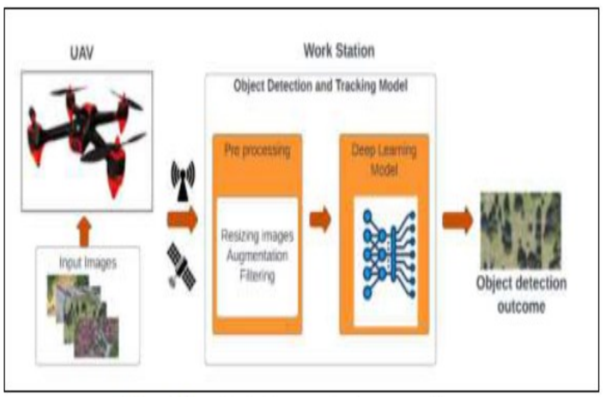

Figure 1 Sample Architecture for drone surveillance system

A great deal of investigation and observation is still necessary, despite the fact that DL does produce encouraging outcomes. Hyperspectral imaging sensors (HIS) have attracted interest for their ability to capture high-resolution images of the earth's surface, which can reveal information about the physical and chemical properties of objects and terrain. This has led to various object detection applications to use these images (Petersson et al., 2017; Signoroni et al., 2019).

III. RESEARCH MOTIVATION

Computer vision, audio recognition, and natural language processing are just a few areas where deep learning techniques have shown great promise for object identification. Algorithms based on deep learning can automatically sift through mountains of data in search of characteristics useful for object recognition. As a result, deep learning is an intriguing strategy for drone surveillance object identification. Limited computational resources, poor data quality, and the need for strong algorithms are just a few of the many obstacles that must be overcome before deep learning can reach its full potential. In order to assess the present state-of-the-art, find research gaps, and suggest future research paths, it is essential to conduct a review of object identification in drone surveillance using deep learning.

IV. RESEARCH CONTRIBUTION

This literature review aims to do just that by surveying the current research on object identification in drone surveillance applications of deep learning. The following will be the review's primary contributions: 1. The purpose of this study is to compare and contrast the object identification performance of different deep learning algorithms and architectures, as well as to highlight their advantages and disadvantages, possible uses, and obstacles. 2. Our goal in doing this evaluation is to shed light on the existing state-of-the-art and suggest avenues for future research that might enhance the efficacy of object detection in drone surveillance. 3. Our goal is to make accessible drone datasets that include the relevant information so that more study may be conducted in this field.

V. THE METHODOLOGICAL FRAMEWORK FOR LITERATURE REVIEW

We anchored the whole literature research procedure on the following questions:
Question 1: How have the most recent advanced object identification algorithms for drone surveillance, which are based on deep learning, changed over the years?
The second question is, how can we get the most out of object identification algorithms that use deep learning for drone surveillance by tweaking their hyperparameters?
Question 3: How can we enhance the effectiveness of object identification algorithms in drone surveillance that rely on deep learning by using transfer learning techniques?
Question 4: How have researchers dealt with the difficulties of training object identification systems based on deep learning for use in drone surveillance? Additionally, what are the necessary directions for the field to progress in the future?

VI. SOLUTIONS TO RESEARCH QUESTIONS

Q1. What are the state-of-the-art deep learning-based object detection algorithms for drone surveillance, and how have they evolved over time?

Although object recognition often made use of older computer vision methods like Haar cascades and HOG (histogram of oriented gradients) before 2014. On the other hand, deep convolutional neural networks (CNNs) like AlexNet—which took first place in 2012's ImageNet Large Scale Visual Recognition Challenge—and other similar architectures began to replace traditional object recognition methods around 2014. During that time, there was a lot of talk about drones having potential uses beyond just surveillance, but few people were utilizing deep learning to identify objects in the sky. Object identification tasks in different UAV applications, including surveillance, were initially underutilized until the following years, when deep learning methods like YOLO, SSD, and Faster R-CNN gained popularity. Since then, deep learning algorithms have seen substantial field applications (Figure 2). In the field of object identification in particular, R-CNN, Faster R-CNN, YOLO, and SSD are among the most popular deep learning techniques. These algorithms have shown state-of-the-art performance on object identification tasks, and they all rely on CNNs for feature extraction. Some more algorithms, such CenterNet, Mask RCNN, M2Det, CPN, and FoveaBox, were presented at the beginning of 2018 and are gradually becoming famous among academics for their applications. There are essentially three types of deep learning algorithms: one-stage, two-stage, and advanced detectors.

a. One Stage Detectors

An example of an object detection approach in deep learning is the one-stage detector, which uses a single neural network pass to directly predict the bounding boxes and class probabilities of objects. It begins by suggesting potential objects or areas of interest, and then it sorts and improves them. A few well-known one-stage detectors include YOLO (You Only Look Once), SSD (Single Shot Detector), and RetinaNet (Redmon et al., 2016; Liu et al., 2016; Lin et al., 2020). YOLO predicts the class probabilities and bounding boxes for each cell in the input picture by splitting it into a grid of cells. A confidence score is connected with each anticipated bounding box that indicates the likelihood that the box includes an item.

b. Two Stage Detectors

While two-stage detectors are a strong family of object detection models, they are computationally costly and very sensitive to the quality of candidate object suggestions; yet, they provide excellent accuracy and versatility. One of the most well-liked two-stage detector designs is the R-CNN family, which comprises the Fast, Faster, and Mask variants of the Region-based Convolutional Neural Network (R-CNN).

In order to provide potential item suggestions, these models often use a Region Proposal Network (RPN), which is then classified by a second network. Some more well-known two-stage detectors include Hybrid Task Cascade, Cascade R-CNN, and Feature Pyramid Network (FPN). In a nutshell, two-stage detectors function by first generating proposals, and then classifying those proposals. The first step is for the model to take the input picture and come up with a list of potential objects to include. The usual tool for the job is a neural network called a Region Proposal Network (RPN), which can be trained with an image and then produces a collection of bounding boxes that could include objects. In most cases, the RPN will generate an input picture feature map using a series of convolutional layers. After that, a tiny window is moved over the feature map and a set of predetermined anchor boxes are applied to each position to create a collection of possible object suggestions.

Stage two involves the model deciding whether each proposed item should be in the forefront (having an object present) or the background (having none).

c. Advanced Detectors.

In comparison to one-stage and two-stage detectors, advanced detectors excel in either efficiency or accuracy when it comes to object detection. Enhanced detectors include EfficientDet, CenterNet, YOLOv4, and DETR, to name a few. Among Google's object detectors, EfficientDet stands out for its ability to achieve state-of-the-art accuracy with a fraction of the parameters and processing required by earlier approaches. For the optimal accuracy-to-efficiency tradeoff, it employs a compound scaling approach to optimize the model's size and depth. A new data augmentation approach called mosaic augmentation, a revamped Darknet backbone network with additional layers, and the usage of anchor boxes with variable aspect ratios and sizes are all part of YOLOv4, the fourth version of YOLO. Table 1 displays a comprehensive comparison of the three detectors.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 1, 2025

Table 1. Comparison of different deep learning object detection techniques based on several performance constraints

Parameters	One Stage Detector	Two Stage Detector	Advanced Detector
Accuracy	Less	Medium	High
Speed	Faster	Slower	Faster
Model size	Smaller	Complex	Optimal
Data Volume	smaller datasets	Require large dataset	Versatile
Object size and shape	small objects	complex object shapes	multi-scale feature fusion
Training time	Less	Longer	Less

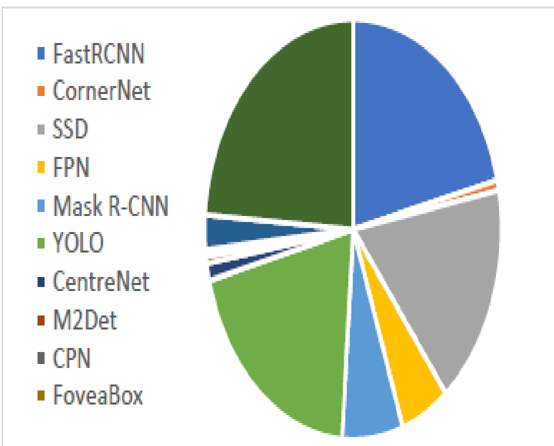

Figure 2 Relative percentage of different deep learning papers published in the UAV domain

Q2. How can we optimize the hyperparameters of deep learning-based object detection algorithms for drone surveillance to achieve optimal performance?

Due to the time-consuming nature of training deep neural networks, optimization is an essential part of deep learning. Deep learning optimizers such as RMSProp, Adagrad, mini batch stochastic gradient descent optimizers, and stochastic gradient descent deep learning optimizer have been developed and used by researchers in various domains (Cui et al., 2018; Shallue et al., 2018; Zhang et al., 2019; Cui et al., 2018; Xu et al., 2021).Nevertheless, there are several methods for optimizing object identification models during execution, such as data augmentation, normalization, transfer learning, neural network learning rate adjustment, feature pyramid networks, and non-maximum suppression.

Data augmentation is a regularization technique that involves enhancing the training data with controlled fluctuations. Overfitting happens when a model becomes too specific to its training data and cannot generalize to new instances; regularization methods assist avoid this. In order to improve the model's object detection accuracy on unseen data and decrease overfitting, data augmentation is used to provide various instances. This enables the model to acquire more robust and generalizable features. In a 2014 study, Girshick et al. Utilizing the R-CNN framework for object identification, the article suggested a two-stage method for object detection, first involving the generation of region suggestions and then categorizing these suggestions using a CNN. Although not directly stated as "data augmentation," the a2uthors used a method of data augmentation during training by randomly scaling and flipping the input pictures horizontally. With this method, the model's resilience and generalizability were both enhanced. Similarly, a number of publications have used data augmentation methods in the preprocessing phase to improve the object identification performance of DL models (Ottoni et al., 2023; Ruiz-Ponce et al., 2023)—just like that. Zhang et al. (2019) found that improving convergence is another way to optimize deep learning models. In order to speed up the training of CNN models, batch normalization approaches were included into the model architecture by Ioffe and Szegedy (2015). These techniques acted as regularizers. The use of normalization allowed us to attain the same accuracy in 14 fewer cycles, eliminating the requirement for dropout. A preprocessing strategy was suggested by Koo and Cha (2017) to improve the performance of an image recognition model. This technique applies normalization to a CNN classifier and feature extraction. The normalized picture is recognized using a calibrated CaffeNet model. With the use of a size-normalized picture, the CNN model was able to improve its performance from an average of 93.24% to 96.85%. When it comes to improving deep learning models for object identification, one more optimization strategy that has been successful is the transfer learning methodology (Aytar, 2014). Transfer learning does this by transferring learned information from one job to another, and it does this by making use of pretrained models on massive datasets.

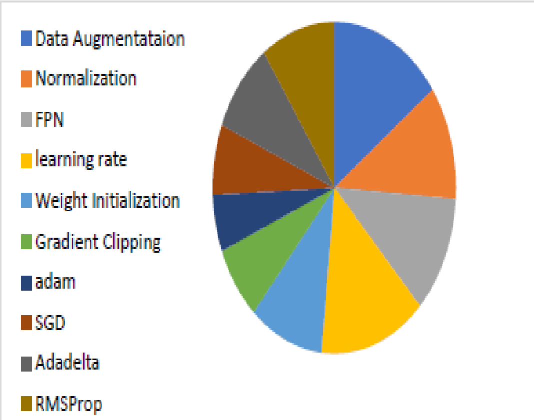

Figure 3 Count of each Optimization techniques applied on deep learning algorithm in various literatures towards object detection

In order to achieve the highest possible detection accuracy, the authors of (Chamarty, 2020) focused on optimizing CNN's learning rate. A correlation between learning rate and dataset size in the 10-4 to 10-5 range was established in the article. In (Na, 2022), a similar strategy was utilized; specifically, a learning rate optimization was implemented, which modifies the learning rate by modifying the approach to multipliers. When compared to existing adaptive gradient algorithms, the suggested strategies for modifying the learning rate performed better. A study conducted by Yang et al. (2022) shown that Feature Pyramid Networks are capable of efficiently handling objects of varying sizes. To increase performance in tasks such as object identification, instance segmentation, semantic segmentation, and Non-Maximum Suppression, FPN combines multiscale information in a feature pyramid. This allows for the detection and recognition of objects of varied sizes (Song et al., 2019). The domain's optimization strategies are detailed in Figure 3.

Q3. How can we leverage transfer learning techniques to improve the performance of deep learning-based object detection algorithms in drone surveillance?

Drone surveillance systems that use deep learning for object identification may greatly benefit from transfer learning. Drone surveillance algorithms that use deep learning for object recognition might benefit from this method's ability to reduce computing resources, speed up training, and enhance performance. Algorithm 1 lays out a comprehensive, step-by-step process for applying transfer learning to object identification, as seen here:

Algorithm 1: Step by step process of transfer learning for object detection
1. Select a pre trained model.
2. Choose the input data extracted through drone.
3. Transfer Learning Process
I. Load the pre-trained model and freeze the early layers
II. Use the pre-trained model as feature extractor
III. Training and fine tuning (Iterate through steps 3(i), (ii), (iii))
i. Train the modified model, update the weights for new layers, retain the knowledge gained from previous steps.
ii. Adjusting parameters such as learning rate, batch size, optimizer, and regularization techniques.
iii. Asses the performance based on precision, recall and f1 score.
iv. Fine-tuning the model or adjusting hyperparameters, include re-annotating data, collecting additional data, or experimenting with different model architectures.
4. Final Output (A fine-tuned or Adapted Model)

Q4. What are the challenges associated with training deep learning-based object detection algorithms for drone surveillance?

Several obstacles arise during the training of object identification systems based on deep learning for use in drone surveillance:
1. Scarcity of labeled data: Capturing and annotating a dataset that encompasses different situations, weather, illumination, and item changes in the drone's field of view demands a substantial investment of time and money. When labeled data is few, it might be difficult to train the model and make it more or less applicable to real-world scenarios.
2. Change in domain: conventional object identification datasets are generally used in conventional object detection systems, however drone surveillance often uses new imaging settings. High altitude, changing views, occlusions, and motion blur are some of the particular issues that drones bring to aerial photography and videography. Pretrained models may struggle to generalize to the domain of drone surveillance due to domain shifts caused by these factors. Additional training or fine-tuning may be necessary if the model has trouble

properly detecting objects in these new settings.
3. Resolution and object size: Depending on the drone's height and distance from the target items, drone surveillance may identify things at different sizes. It might be somewhat tough for the model to recognize and locate objects effectively in scenes when their size varies significantly or when they look tiny.

Furthermore, the quality and visibility of the objects in the recorded movies or photographs may be compromised due to the low resolution of drone cameras. If you want accurate object identification results, you have to solve these problems with size and resolution.

To make quick decisions, drone surveillance applications often need object detection in real-time or near real-time. This brings us to our fourth point: real-time performance. Since drones often have limited computing power, it might be difficult to implement deep learning-based object identification algorithms at the required speed. Improving efficiency in real-time while maintaining accurate detection may need optimization methods such as model reduction, quantization, or hardware acceleration.

5. Adjusting to ever-changing environments: Capturing images with moving objects and shifting backdrops is a common part of drone surveillance. There could be intricate patterns of motion, occlusions, or interactions involving the items of interest. A diversified dataset including multiple motion patterns and item interactions is necessary for training a model that can handle such dynamic scenarios efficiently. In order to extract the time-related data from the drone surveillance footage, meticulous model architecture and temporal modeling approaches are required.

6. Drones' restricted flight duration is caused by their battery capacity, which in turn limits the quantity of data that can be acquired during each flying session. Because of this restriction, collecting a big enough and representative dataset is not an easy task. Further restricting the dataset's variety and quantity are operational limits, privacy issues, and rules that may ban data collecting in some places or under specified situations.

VII. CONCLUSION

Despite continued efforts to dispel this myth, deep learning (DL) is still often seen as a "black-box" answer to many issues. Deep learning (DL) has already significantly advanced many areas of remote sensing. The application of DL approaches to the analysis of photos taken by UAVs is the primary emphasis of our literature study. In order to achieve this goal, our research will provide a synopsis of current methods and viewpoints on how to use them in order to give a thorough grasp of the topic. We want to provide a comprehensive overview of the uses of DL-based methods in UAV image processing via this literature review. It is determined from this review that:

1. While most published works on object recognition using deep learning focus on convolutional neural networks (CNNs) and radial basis functions (RCNNs), multi-and hyperspectral data might be useful in some applications, such as precision agriculture and forest-related tasks.

There is an obvious need for more publicly accessible datasets particularly acquired with UAVs to improve network training and benchmarking. If researchers want to use supervised learning techniques and successfully train and assess their networks, they need to make sure these datasets are appropriately labeled.

3. Fast and efficient data processing is made possible by combining deep learning (DL) methods with GPU processing, which allows for fast inference solutions. More investigation into real-time processing using embedded systems developed for UAVs is still required, however.

VIII. REFERENCES

[1]. Aytar, Y. (2014). Transfer learning for object category detection. PhD Thesis, 146. https://www.robots.ox.ac.uk/~vgg/publications/2014/Aytar14a/aytar14a.pdf

[2]. Chamarty, A. (2020). Fine-Tuning of Learning Rate for Improvement of Object Detection Accuracy. Proceedings - 2020 IEEE India Council International Subsections Conference, INDISCON 2020, 135–141. https://doi.org/10.1109/INDISCON50162.2020.00038

[3]. Cui, X., Zhang, W., Tüske, Z., & Picheny, M. (2018). Evolutionary stochastic gradient descent for optimization of deep neural networks. Advances in Neural Information Processing Systems, 2018-Decem(NeurIPS), 6048–6058.

[4]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81

[5]. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456.

[6]. Koo, K. M., & Cha, E. Y. (2017). Image recognition performance enhancements using image normalization. Human-Centric Computing and Information Sciences, 7(1). https://doi.org/10.1186/s13673-017-0114-5

[7]. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

[8]. Liu, W., Anguelov, D., Erhan, D., & Fu, C.-Y. (2016). SSD:Single Shot MultiBox Detector. European Conference on Computer Vision, 1(January 2017), 398–413. https://doi.org/10.1007/978-3-319-46448-0_2

[9]. Long, J., Darrell, T., & Berkeley, U. (2015). Fully Convolutional Networks for Semantic Segmentation. CVPR, IEEE Access, 7, 3431–3440. https://doi.org/10.1109/ACCESS.2019.2908685

[10]. Na, G. S. (2022). Efficient learning rate adaptation based on hierarchical optimization approach. Neural Networks, 150, 326–335. https://doi.org/10.1016/j.neunet.2022.02.014

[11]. Ottoni, A. L. C., de Amorim, R. M., Novo, M. S., & Costa, D. B. (2023). Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. International Journal of Machine Learning and Cybernetics, 14(1), 171–186. https://doi.org/10.1007/s13042-022-01555-1

[12]. Petersson, H., Gustafsson, D., & Bergström, D. (2017). Hyperspectral image analysis using deep learning - A review. 2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA 2016. https://doi.org/10.1109/IPTA.2016.7820963

[13]. Redmon, J., Santhosh, D., Ross, G., & Farhadi, A. (2016). You Only Look Once: Unified , Real-time Object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitio, 779–788. https://doi.org/10.1145/3243394.3243692

[14]. Ruiz-Ponce, P., Ortiz-Perez, D., Garcia-Rodriguez, J., & Kiefer, B. (2023). POSEIDON: A Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments. Sensors (Basel, Switzerland), 23(7), 1–14. https://doi.org/10.3390/s23073691

[15]. Shallue, C. J., Lee, J., Antogniniy, J., Sohl-Dickstein, J., Frostig, R., & Dahl, G. E. (2018). Measuring the effects of data parallelism on neural network training. ArXiv, 20, 1–49.

[16]. Signoroni, A., Savardi, M., Baronio, A., & Benini, S. (2019). Deep learning meets hyperspectral image analysis: A multidisciplinary review. Journal of Imaging, 5(5). https://doi.org/10.3390/jimaging5050052

[17]. Song, Y., Pan, Q. K., Gao, L., & Zhang, B. (2019). Improved non-maximum suppression for object detection using harmony search algorithm. Applied Soft Computing Journal, 81, 105478. https://doi.org/10.1016/j.asoc.2019.05.005

[18]. Xu, D., Zhang, S., Zhang, H., & Mandic, D. P. (2021). Convergence of the RMSProp deep learning method with penalty for nonconvex optimization. Neural Networks, 139, 17–23. https://doi.org/10.1016/j.neunet.2021.02.011

[19]. Yang, X., Wang, W., Wu, J., Ding, C., Ma, S., & Hou, Z. (2022). MLA-Net: Feature Pyramid Network with Multi-Level Local Attention for Object Detection. Mathematics, 10(24), 1–13. https://doi.org/10.3390/math10244789

[20]. Zhang, N., Lei, D., & Zhao, J. F. (2019). An Improved Adagrad Gradient Descent Optimization Algorithm. Proceedings 2018 Chinese Automation Congress, CAC 2018, 3, 2359–2362. https://doi.org/10.1109/CAC.2018.8623271